RF-GAN: A Light and Reconfigurable Network for Unpaired Image-to-Image Translation

Ali Köksal[1,2] and Shijian Lu[2]

[1] Institute for Infocomm Research, A*Star, Singapore
[2] Nanyang Technological University, Singapore

Abstract. Generative adversarial networks (GANs) have been widely studied for unpaired image-to-image translation in recent years. On the other hand, state-of-the-art translation GANs are often constrained by large model sizes and inflexibility in translating across various domains. Inspired by the observation that the mappings between two domains are often approximately invertible, we design an innovative reconfigurable GAN (RF-GAN) that has a small size but is versatile in high-fidelity image translation either across two domains or among multiple domains. One unique feature of RF-GAN lies with its single generator which is reconfigurable and can perform bidirectional image translations by swapping its parameters. In addition, a multi-domain discriminator is designed which allows joint discrimination of original and translated samples in multiple domains. Experiments over eight unpaired image translation datasets (on various tasks such as object transfiguration, season transfer, and painters' style transfer, etc.) show that RF-GAN reduces the model size by up to 75% as compared with state-of-the-art translation GANs but produces superior image translation performance with lower Fréchet Inception Distance consistently.

1 Introduction

Image-to-image translation aims to translate images from a source domain to a target domain so that the translated images have similar appearance, styles, etc. as the images in the target domain. With the fast development of generative adversarial networks (GANs), quite a number of GANs have been reported in recent years which are capable of generating very realistic image-to-image translations in terms of object appearance [1–6], painting styles [7–12], seasonal styles [13, 14], etc.

Image-to-image translation GANs can be broadly classified into two categories according to their scalability. The first category performs image translation across two domains only which typically involve two translators (each consists of a generator and a discriminator) such as CycleGAN [14], DiscoGAN [3], and UNIT [13]. The second category performs image translation across multiple domains which typically employs a single generator, a single discriminator as well as additional classifiers (for handling multi-domain translation) such as StarGAN [15] and DosGAN [16]. These translation GANs have a common constraint

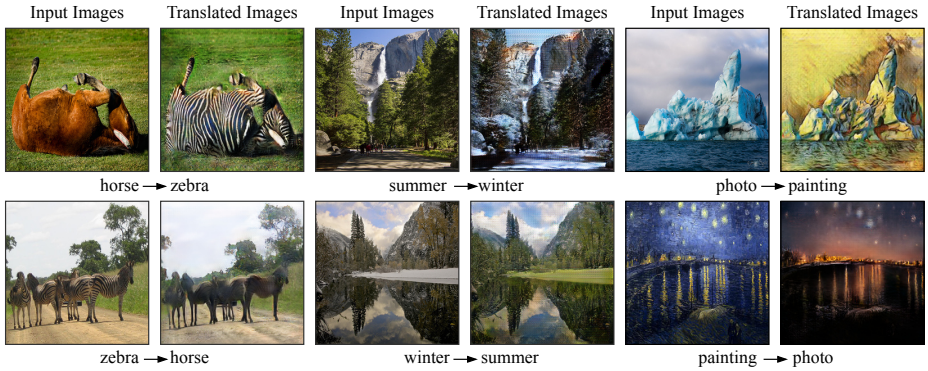

Fig. 1. The proposed RF-GAN learns a single generator for image translations in opposite directions. For translation on object transfiguration, season transfer, and painter style transfer from left to right, RF-GAN can translate images with a reconfigurable G as shown in the top row. By simply reconfiguring G to a new generator G_r via parameter swapping, RF-GAN can translate images back in the opposite direction as shown in the bottom row. RF-GAN reduces the model size by up to 75% as compared with state-of-the-art GANs such as CycleGAN and StarGAN but obtains overall lower Fréchet inception distance (FID) over eight unpaired image translation datasets.

that they usually involve a large number of network parameters either due to the double-generator-double-discriminator architecture in cross-domain translation GANs or the additional classifiers in multi-domain translation GANs. As a result, they often face various limitations in many resource-constrained scenarios such as edge computing. Additionally, they also require a large amount of images for training high-fidelity translation models due to the large amount of parameters involved. Another common constraint is about the limited flexibility. Specifically, cross-domain translation GANs cannot scale to handle multi-domain translation tasks without an increase in model size. Multi-domain translation GANs can handle cross-domain translation but their performance often drops a lot as their classifiers encourage generator for multi-domain translation and are susceptible to the noise added by generators.

This paper presents an innovative reconfigurable GAN (RF-GAN) that is small with just a single translator but capable of performing high-fidelity image translation across two domains or among multiple domains. RF-GAN is designed based on the observation that bidirectional mappings between domains are often approximately invertible. It learns a single translator for bidirectional image translations, where the forwards and backwards translations are achieved by swapping the parameters of the same generator as illustrated in Figs. 1 and 2. In addition, a multi-domain discriminator is designed which can discriminate images in opposite translation directions without either multiple domain-specific discriminators or additional classifier as required by most existing translation

GANs. Further, RF-GAN can be trained with less training images, or better trained with the same amount of training images as state-of-the-art translation GANs. This is partially due to the much fewer network parameters in RF-GAN (up to 75% less than state-of-the-art GANs as the reconfigurable generators G and G_r share the same set of parameters) that require less images to train. Additionally, RF-GAN employs a single discriminator only which achieve similar effect of doubling training data as compared with state-of-the-art GANs that employ two discriminators or extra classifiers.

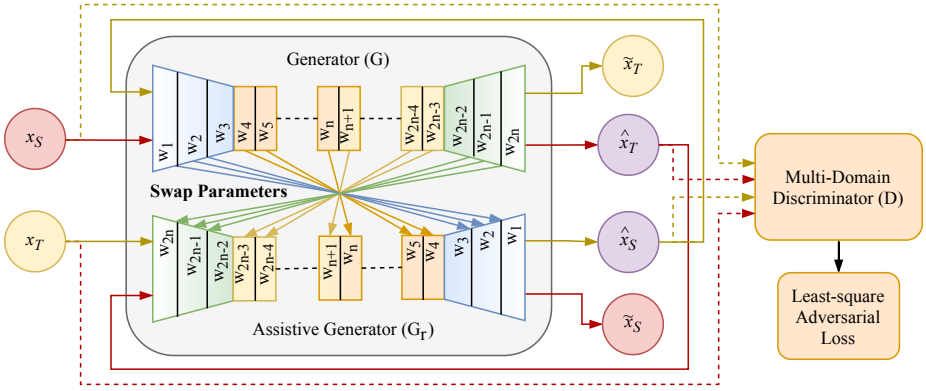

Fig. 2. The architecture of the proposed RF-GAN: RF-GAN consists of a reconfigurable generator G and a multi-domain discriminator D. In each training iteration, G first learns to translate images x_S in domain S to images $\hat{x}_T$ in domain T. G is then reconfigured to an assistive generator G_r by swapping its parameters which learns to translate $\hat{x}_T$ to $\tilde{x}_S$ as well as x_T to $\hat{x}_S$. After that, G_r is reconfigured back to G which further learns to translate from x_S to $\hat{x}_T$ as well as $\hat{x}_S$ to $\tilde{x}_T$ ($\tilde{x}_S$ and $\tilde{x}_T$ for computing reconstruction loss). The multi-domain discriminator D is trained continuously after each translation with x_S, $\hat{x}_T$, x_T, and $\hat{x}_S$ to compute least-square adversarial loss.

The contributions of this work can be summarized in three aspects. First, it designs an innovative RF-GAN that is capable of performing high-fidelity image translation across two domains or among multiple domains. Second, it designs a reconfigurable generator G and a multi-domain discriminator D where G can achieve bidirectional image translation by swapping its parameters and D can perform multi-domain translation without requiring multiple domain-specific discriminators or additional classifiers. As a result, RF-GAN reduces the model size by up to 75% as compared with state-of-the-art GANs, and it can be better trained with the same amount of training images. Third, extensive experiments show that RF-GAN outperforms state-of-the-art GANs such as cross-domain CycleGAN and multi-domain StarGAN consistently across eight public datasets.

2 Related Work

2.1 Generative Adversarial Networks (GANs)

The idea of the original Generative Adversarial Networks (GANs) [17] is to train a generator and a discriminator in an adversarial manner. Specifically, the generator is trained to generate images as realistic as possible for fooling the discriminator, the discriminator is instead trained to distinguish the generated images from real ones as accurate as possible. GANs trained by such adversarial learning can often generate very impressive and realistic images.

With the great success of the adversarial learning, GANs have been studied extensively in recent years [18–21] with applications in various tasks such as image inpainting [22], image synthesis [23–25, 20, 26, 27], video generation [28], and 3D modelling [29]. They have also been widely studied for image-to-image translation, more details to be described in the next subsection.

2.2 Image-to-image Translation

Image-to-image translation aims to transform images from one domain to another where images often have different characteristics such as colors, styles, etc. Quite a number of GANs have been designed for the task of image-to-image translation in recent years [2, 9, 15, 30–38], starting from earlier methods that require paired training images from different domains to the recent that can be trained with unpaired images.

For GANs requiring paired training images, [2] presents Pix2pix, a general-purpose translation network that adopts conditional GAN (cGAN) [25] to learn mappings between two sets of paired images. Similarly, [39] employs cGAN to learn the mapping from sketches to photos and AL-GAN [40] uses cGAN to generate scene images conditioning on scene attributes and layout. In addition, [41] presents a $\triangle$-GAN that uses semi-supervised learning for cross-domain joint distribution matching. To address the lack of diversity of the aforementioned methods, BicycleGAN [42] is designed to generate continuous and multimodal distribution. Paired images provide useful supervision information but collecting paired images from different domains is often time-consuming.

For GANs that can work with unpaired training images, DTN [6] introduces a network with one generator and one discriminator for general purpose translation. Co-GAN [4] proposes a two-generator-two-discriminator network that learns joint distribution of multi-domain images by sharing a latent space. Similarly, UNIT [13] uses a shared latent space but involves a complex framework with two encoders, two generators, and two discriminators. CycleGAN [14] employs two generators to learn bidirectional mappings between two domains and it also employs two discriminators for each of the two domains. Similar to CycleGAN, DiscoGAN [3] uses two generators and two discriminators and employs cycle consistency and reconstruction losses to measure how well source domain images are translated back after translating to the target domain. Similar to BicycleGAN, MUNIT [43], and DRIT [44] aims to generate multiple outputs from one input by decomposing images as content and style.

To address the lack of generalization, many existing cross-domain translation GANs can be adapted to multi-domain translation tasks. In the straightforward adaptation, one model can be used for each binary combination of domains and trained separately. For example, ComboGAN [30] extends cross-domain translation to multi-domain by training less model than straightforward adaptation. In addition, some efforts aim for multi-domain image translation with a small model. For example, StarGAN [15] proposes a model that contains a single conditioned generator, a single discriminator and an auxiliary classifier on top of the discriminator. DosGAN [16] similarly employs a single generator, a single discriminator, and a single pre-trained classifier. SingleGAN [37] instead employs a single generator but multiple discriminators.

Cross-domain translation GANs such as DiscoGAN and CycleGAN have two generators and two discriminators that introduce a large number of network parameters which require a large amount of training images to train. Multi-domain translation GANs such as StarGAN and DosGAN employ a single generator and a single discriminator, but its discriminator is a large network. Our proposed RF-GAN has a single generator and a single discriminator without additional classifier which reduces up to 75% network parameters by comparing the translation GANs. This expands the RF-GAN's applicability in resource-constrained devices and also helps reduce the required training images greatly. On the other hand, it achieves superior translation fidelity consistently as compared with most state-of-the-art translation GANs, largely due to the reconfigurable generator and the multi-domain discriminator to be described in the following Section.

3 The Proposed Method

The proposed RF-GAN learns a single generator for image mapping in opposite directions between domains. It also learns a multi-domain discriminator for bidirectional image discrimination without requiring domain-specific discriminators or extra classifier. A novel training strategy is designed to train the proposed RF-GAN effectively, more details to be presented in the following subsections.

3.1 Reconfigurable Generator

Learning a mapping function to map images in opposite directions takes an iterative learning process in the proposed RF-GAN. Fig. 2 shows one learning iteration, where the generator G and G_r (derived by swapping $G's$ parameters) both have an encoder and a decoder. Given images x_S from a source domain S, G first learns to map them to $\hat{x}_T$ in a target domain T conditioned to the category of the target domain. Once G is learned, it is reconfigured to an assistive generator G_r automatically by swapping *G's* parameters. G_r then learns the mapping from x_T to $\hat{x}_S$ as well as from $\hat{x}_T$ to $\tilde{x}_S$. After that, G_r is reconfigured back to G by swapping its parameters which will further learn the mapping from x_S to x_T and from $\hat{x}_S$ to $\tilde{x}_T$. The iterative learning will finally lead to a single reconfigurable generator G. Given new images from domains S and T, G can map images from

S to T, and it can simply be reconfigured to G_r (by swapping its parameters) for inverse mapping from T to S. Details of the parameter swapping and generator training will be described in the following three subsections.

Parameter Swapping The target of the proposed parameter swapping is to learn one generator for image translations in opposite directions. Given images in domains S and T, a mapping function G will be confused and fail to learn well if we directly train it for mappings in the directions $S \rightarrow T$ and $T \rightarrow S$ concurrently. The reason is that for the mapping $S \rightarrow T$, *G's* encoder parameters will mainly deal with images in domain S and *G's* decoder parameters will mainly deal with images in domain T. If we concurrently train G for the mapping $T \rightarrow S$, *G's* encoder parameters will have to deal with images in domain T and its decoder parameters will have to deal with images in domain S. This will confuse G and lead to an undesired mapping function.

We introduce an assistive generator G_r to learn the mapping $T \rightarrow S$. To ensure that we will finally learn a single generator, we design G and G_r in a way that G_r can be simply derived from G by swapping *G's* parameters. In this way, the proposed RF-GAN first learns G for the mapping $S \rightarrow T$ and then derives G_r (automatically by swapping *G's* parameters) for learning the mapping $T \rightarrow S$. The iterative learning leads to a reconfigurable G that can handle image translations in opposite directions. For a $2n-$layer network, the mappings for G and G_r can be formulated as follows:

$$\begin{aligned} \mathbb{E}_{x_S \sim X_S} G(x_S, y_T) =& \hat{x}_T \quad \text{, where} \\ G(x_S, y_T) =& x_S \odot W_1 \odot W_2 \odot \ldots \odot W_n \odot W_{n+1} \odot \ldots \odot W_{2n} \end{aligned} \tag{1}$$

$$\begin{aligned} \mathbb{E}_{x_T \sim X_T} G_r(x_T, y_S) =& \hat{x}_S \quad \text{, where} \\ G_r(x_T, y_S) =& x_T \odot W_{2n} \odot W_{2n-1} \odot \ldots \odot W_{n+1} \odot W_n \odot \ldots \odot W_1 \end{aligned} \tag{2}$$

where $\odot$ denotes convolution operations, y_S and y_T are the category of domains source S and target T, respectively, and $W_1, ..., W_{2n}$ denote the convolutional layers as shown in Fig. 2.

Generator Architecture Leveraging on the CycleGAN generator [14], we design a reconfigurable generator for mapping images in opposite directions. Our generator replaces all convolution layers with fractional-strided convolution layers in the second half of the architecture of CycleGAN generator. For a detailed comparison. As fractional-strided convolution is widely used for deconvolution, the use of convolution layers in the first half and fractional-strided convolution layers in the second makes the reconfigurable generator invertible and more suitable for learning bidirectional mapping. Due to the symmetric structure of the reconfigurable generator, convolution layers and fractional-strided convolution layers have the same number of parameters. In another word, W_{1+i} and W_{2n-i},

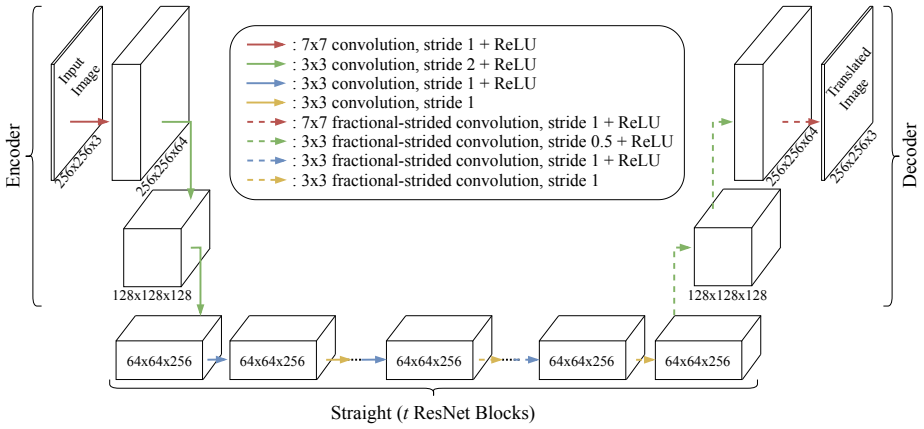

Fig. 3. The proposed reconfigurable generator G has perfect symmetric structures and so symmetric input dimensions which ensure that the parameter swapping in G can be carried out without discrepancy.

$i \in [0, n)$ have exactly the same size. This symmetric structure also ensures that the input of each layer (activation size in each layer) has the same symmetric relation as shown in Fig.3. Note the parameters of CycleGAN generator can also be swapped with certain adaptations such as parameter transposition, but the performance of the new generator is much lower than ours that has a more invertible structure.

As shown in Fig. 3, the architecture consists of three major components including an encoder that progressively down-samples by two convolution with stride 2, a decoder that progressively up-samples by two fractional-strided convolution with stride 0.5, and a straight component with t ResNet blocks [45] (default at 9) in the middle. Similar to the CycleGAN generator, the first half of the straight component in our generator is convolution layers with stride 1. However, the second half changes to fractional-strided convolution layers with stride 1 for parameter swapping. In addition, our generator uses a fractional-strided convolution layer with the same stride as the last convolution layer. Further, both CycleGAN generator and our generator use reflection padding in the first half of the straight component and down-sampling layers, but our generator uses inverse padding (which crops activations from the edges of activations symmetrically) in the second half of the straight and up-sampling layers. Note normalization layers are omitted for the sake of the visual simplicity in Fig. 3.

Generator Training The generators G and G_r learn alternatively while training the RF-GAN iteratively. In each training iteration, G first translates image x_S in source domain S to images $\hat{x}_T$ in target domain T. After that, G is reconfigured to G_r by automatic swapping G's parameters and G_r then learns to translate $\hat{x}_T$ to $\tilde{x}_S$. Intuitively, x_S and $\tilde{x}_S$ should be the same, but they are

never perfectly the same as G_r is not a perfect inverse of G. A reconstruction loss [3], [14] between x_S and $\tilde{x}_S$ should therefore be computed to train G and G_r to learn the mappings in the two opposite directions and force them to be approximately inverse of each other:

$$\begin{aligned}\mathcal{L}^{REC}(G, X_S, X_T) =& \mathbb{E}_{(x_S, y_S)\sim(X_S, Y_S)} \|G_r(G(x_S, y_T), y_S) - x_S\|_2 \\ &+ \mathbb{E}_{(x_T, y_T)\sim(X_T, Y_T)} \|G(G_r(x_T, y_S), y_T) - x_T\|_2\end{aligned} \tag{3}$$

In addition, a least-square adversarial loss [46] should be computed with translated images as follows:

$$\begin{aligned}\mathcal{L}_G^{GAN}(G, D, X_S, X_T) =& \mathbb{E}_{(x_S, y_S)\sim(X_S, Y_S)}(D(G(x_S, y_T)) - y_T)^2 \\ &+ \mathbb{E}_{(x_T, y_T)\sim(X_T, Y_T)}(D(G_r(x_T, y_S)) - y_S)^2,\end{aligned} \tag{4}$$

where y_S is the label of images in S (x_S) and y_T is the label of images in T (x_T).

The training of G should minimize both reconstruction loss and adversarial loss as formulated as follows:

$$\mathcal{L}_G(G, D, X_S, X_T) = \mathcal{L}_G^{GAN}(G, D, X_S, X_T) + \lambda \mathcal{L}^{REC}(G, X_S, X_T) \ , \tag{5}$$

where λ controls the relative effect between adversarial and reconstruction losses.

Fig. 4. For images x_S in domain S, our reconfigurable generator G can translate them to $\hat{x}_T$ in domain T which can be translated back to $\tilde{x}_S$ in domain S by generator G_r that can be simply reconfigured from G.

Fig. 4 illustrates the image mapping with our proposed reconfigurable generator G. As Fig. 4 shows, the reconfigurable generator G is capable of translating images x_S in domain S to images $\hat{x}_T$ in domain T as shown in columns 1 & 4 and columns 2 & 5. At the same time, the reconfigured G_r (from G) is capable of translating ($\hat{x}_T$) back to ($\tilde{x}_S$) in domain S as shown in columns 3 & 6. This clearly shows the effectiveness of our proposed RF-GAN that can learn one mapping function for image mapping in two opposite directions.

3.2 Multi-domain Discriminator

Discriminator in GANs is basically a domain-specific binary classifier which aims to distinguish real and translated images in one specific domain. Therefore, translation GANs such as DiscoGAN [3] and CycleGAN [14] employ two

domain-specific discriminators for discriminating images translated in two opposite directions. In the more recent multi-domain translation GANs such as StarGAN [15], a single discriminator is used for multi-domain discrimination but an auxiliary classifier is employed on top of the discriminator for classifying samples according to their domains. This also applies to DosGAN [16] which also uses a pre-trained classifier together with a discriminator for differentiating samples from more than one domain.

We design a multi-domain discriminator D that can perform image discrimination in opposite directions without requiring either more than one domain-specific discriminators or additional classification. Specifically, our multi-domain discriminator learns to discriminate real images in domain T and the translated images from domain S to domain T (by G), as well as the real images in domain S and the translated images from domain T to domain S (by G_r). The training aims to minimize the following adversarial loss:

$$\mathcal{L}_D(G, D, X_S, X_T) = \mathcal{L}_D^{GAN}(G, D, X_S, X_T) + \mathcal{L}_D^{GAN}(G_r, D, X_T, X_S) \tag{6}$$

where the two least-square adversarial loss components [46] can be computed in a similar way. One of them can be computed by:

$$\begin{aligned} \mathcal{L}_D^{GAN}(G, D, X_S, X_T) =& \mathbb{E}_{(x_T, y_T)\sim(X_T, Y_T)}(D(x_T) - y_T)^2 \\ &+ \mathbb{E}_{(x_S)\sim(X_S)}(D(G(x_S, y_T)) - y_T^*)^2 \end{aligned} \tag{7}$$

where y_S, y_T, y_S^*, and y_T^* refer to the label of images in domains S (x_S) and T (x_T), as well as the translated images in domains S ($\hat{x}_S$) and T ($\hat{x}_T$), respectively.

Our multi-domain discriminator is a PatchGAN [47] adapted from [14]. It consists of five 4x4 convolution layers three of which are with stride 2 and the other two are with stride 1. Leaky ReLU is used as the activation function.

Algorithm 1 RF-GAN Training

Input: Generator G, discriminator D, batch of training sets $(x_S, y_S) \in (X_S, Y_S)$ and $(x_T, y_T) \in (X_T, Y_T)$

Output: Updated generator G and discriminator D

1: $\hat{x}_T \leftarrow G(x_S, y_T)$
2: **update** D by $\mathcal{L}_D^{GAN}$ based on x_T and $\hat{x}_T$ with Equation 7
3: $G_r \leftarrow$ reconfigure G
4: $\tilde{x}_S \leftarrow G_r(\hat{x}_T, y_S)$ and $\hat{x}_S \leftarrow G_r(x_T, y_S)$
5: **update** D by $\mathcal{L}_D^{GAN}$ based on x_S and $\hat{x}_S$ with Equation 7
6: $G \leftarrow$ reconfigure G_r
7: $\tilde{x}_T \leftarrow G(\hat{x}_S, y_T)$
8: **update** G by L_G^{GAN} with Equation 4 and L^{REC} with Equation 3

3.3 RF-GAN Training

The RF-GAN employs an assistive generator G_r during its iterative training as illustrated in Fig. 2. Each training iteration consists of two cycles: 1) images x_S in domain S are first translated to $\hat{x}_T$ in domain T by G and then translated back to $\tilde{x}_S$ by G_r (reconfigured from G); 2) images x_T in domain T are translated to $\hat{x}_S$ in domain S by G_r and then translated back to $\tilde{x}_T$ by G (reconfigured from G_r). During this iterative training process, it is critical to keep the association between image domains and the generator G/G_r consistently. Specifically, while G is translating images from domain S to domain T, G_r must translate images from domain T to domain S. Without this association, the cycle consistency will be broken and the generators will be confused easily. We design a sequential training strategy to guarantee this association.

As shown in Algorithm 1 which illustrates single training iteration, generators G and G_r swap their parameters twice to keep their association with image domains. In addition, the update of G is postponed to the end of the training as the loss computation requires the reconstructed images $\tilde{x}_S$ and $\tilde{x}_T$. Different from G, the multi-domain discriminator D is updated twice instead. The first update happens after D learns to discriminate images in domain T x_T and the translated images to domain T $\hat{x}_T$, and the second happens after D learns to discriminate images in domain S x_S and the translated images to domain S $\hat{x}_S$.

During the iterative training of RF-GAN, λ in Eq. 5 weights the reconstruction loss and adversarial loss which is experimentally set at 10. In all evaluations, networks are trained from scratch by applying Adam problem solver [48] with learning rate of 0.0002 and betas 0.5 and 0.999.

4 Experiments

4.1 Datasets and Evaluation Metrics

Datasets: We evaluated RF-GAN over eight public datasets on unpaired image translation. Each dataset consists of a training set and a test set and the eight datasets can be grouped into three categories on object transfiguration, season transfer, and painters' style transfer. The *object transfiguration* category has two datasets including apple↔orange and horse↔zebra. The *season transfer* also contains two datasets including summer↔winter photos of Yosemite and different (four) seasons of the Alps. The *painters' style transfer* has four datasets where each consists of natural photographs in one domain and paintings of one of four artists (Cezanne, Monet, Ukiyo-e, and Van Gogh) in another domain.

Evaluation Metrics: Quantitative evaluation of GAN-synthesized images is still a challenging task [49], [50]. We performed quantitative evaluations by using Fréchet Inception Distance (FID) [51] which is one of the most widely used metrics in evaluating GAN-synthesized images. FID measures the similarity between two sets of samples with the range of $[0, \infty)$. It uses inception model [52] and measures the distance between multivariate Gaussian distribution of features

Methods	DiscoGAN [3]	CycleGAN [14]	StarGAN [15]	RF-GAN
# generator/s	2	2	1	1
# discriminator/s	2	2	1	1
# parameters (M)	16.560	28.286	53.208	14.143

Table 1. Comparison of RF-GAN with DiscoGAN [3], CycleGAN [14], and StarGAN [15] in the number of generators and discriminators as well as parameter numbers.

extracted from an intermediate layer of inception net. While compared with natural images, a lower FID means high fidelity of the synthesized images.

We compare RF-GAN with three state-of-the-art translation GANs including DiscoGAN, CycleGAN, and StarGAN. Table 1 provides their comparison in terms of the number of generator/s, discriminator/s, and size of network parameters with having default number of ResNet block in generators. As Table 1 shows, RF-GAN is 50% smaller than CycleGAN. It is also smaller than DiscoGAN which uses an encoder and decoder with no straight components. Although StarGAN uses a single generator and discriminator, it has many more parameters than RF-GAN because of its auxiliary classifier.

4.2 Experimental Results

In the evaluations, CycleGAN uses a pre-trained model whereas DiscoGAN and StarGAN are trained from scratch. In all evaluations, RF-GAN is trained in the same manner as CycleGAN for 200 epochs from scratch for fair comparisons. Once trained, each sample image in the test set is translated from the source domain to the target domain. FID is then computed between the full test set of the target domain and the generated samples to measure their similarity. Table 2 shows the experimental results where the last column shows the FID between the training and the test sets. Since the training and test sets of photo→Van Gogh contain the same sample images, its FID is zero.

RF-GAN(G) and RF-GAN(G_r) in Table 2 refer to the RF-GANs whose generators are initially trained in the same and opposite directions as listed in Translations column. We can observe that RF-GAN initially trained in either direction achieves similar translation performance, demonstrating the effectiveness of our proposed reconfigurable generator. Note the first four tasks in the RF-GAN column are replicated two times (including RF-GAN training and evaluation) and then compute FID average and variation, just to show the stability of RF-GAN in image translation. The last row shows overall FID scores each of which is computed by the average of normalized FIDs across the ten translation tasks (normalization is computed by Real's FID/GAN's FID). Since the FID of translated images is almost always higher than the FID of real images, the normalized FID usually lies [0, 1] (the bigger the better).

As Table 2 shows, RF-GAN outperforms DiscoGAN consistently by large margins across the ten studied translation tasks. DiscoGAN's much higher FIDs

Translations	Methods					Real
	DiscoGAN [3]	CycleGAN [14]	StarGAN [15]	RF-GAN (G)	RF-GAN (G_r)	*training vs test*
apple→orange	377.58	181.34	222.71	**173.13±1.34**	178.58	55.31
orange→apple	345.54	164.46	167.68	**136.64±0.07**	143.95	48.36
horse→zebra	414.28	81.25	150.39	38.23±0.29	**37.66**	29.62
zebra→horse	333.23	**143.03**	206.15	143.77±1.84	144.05	89.99
summer→winter	296.75	**82.30**	131.11	99.86	101.79	64.64
winter→summer	296.99	**80.11**	120.65	98.70	91.97	44.87
photo→Cezanne	360.41	216.76	280.43	**212.70**	216.14	180.51
photo→Monet	279.41	133.08	172.02	**128.20**	128.82	108.09
photo→Ukiyo-e	316.84	180.33	212.31	166.38	**150.27**	126.09
photo→Van Gogh	322.99	109.77	211.44	**108.69**	109.61	0*
Overall Score	0.25	0.59	0.43	**0.63**	0.63	1

Table 2. Comparison of RF-GAN with state-of-the-art GANs (in FID): RF-GAN(G)/RF-GAN(G_r) denotes our RF-GAN whose generator is initially trained in the same/opposite direction as listed in the column Translations. For object transfiguration and season transfer in the first three tasks, the translations are bidirectional. For the four painters' style transfer tasks, the translations are from photographs to paintings as translating natural photographs to paintings is more meaningful. FID in the last column is computed between real paintings in the training and test sets (* denotes that training and test samples of photo→Van Gogh are the same so FID zero). The last row 'Overall Score' is the average of normalized FID across the nine translation tasks (photo→Van Gogh not included), where the normalized FID for each task is computed by Real's FID/GAN's FID. Note the first four tasks in the RF-GAN column are replicated two times (for RF-GAN training and evaluation) and then compute FID average and variation, just to show the stability of RF-GAN in image translation.

Translations	Methods		Real
	StarGAN [15]	RF-GAN (G)	*training vs test*
(summer+autumn+winter)→spring	106.86	**104.45**	94.95
(spring+autumn+winter)→summer	105.88	**90.92**	73.56
(spring+summer+winter)→autumn	108.66	**96.08**	76.48
(spring+summer+autumn)→winter	101.831	**87.68**	78.16
Overall Score	0.76	**0.85**	1

Table 3. Comparison of RF-GAN with StarGAN (in FID) over different seasons of the Alps dataset (multiple domains). Translated images are generated from test samples of the other seasons. 'Overall Score' is calculated similarly as in Table 2.

could be due to its very simple network structures, though it is still bigger than RF-GAN due to its two-generator-two-discriminator design. In addition, RF-GAN outperforms CycleGAN for object transfiguration and painters' style transfer tasks consistently, though its size is just 50% of the CycleGAN. Further, Tables 2 and 3 show that RF-GAN translates better than StarGAN for either

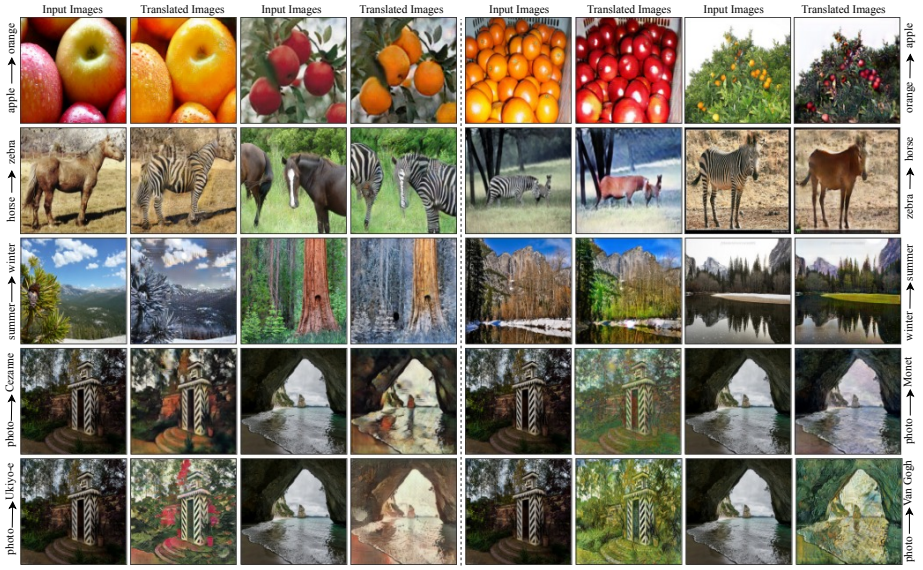

Fig. 5. Illustration of input and translated images by RF-GAN for apple↔orange, horse↔zebra, summer↔winter, photo→Cezanne, photo→Monet, photo→Ukiyo-e, and photo→Van Gogh.

cross-domain or multi-domain translations. Although both have a single generator and discriminator, the parameter number of StarGAN is up to 3.5 times more than RF-GAN. As a result, StarGAN requires more training samples for training a good translator. Besides, translated images by StarGAN tend to be similar to the input images due to the conflict between its generator and auxiliary classifier as discussed in [37]. All these quantitative results demonstrate the superior performance of our proposed RF-GAN.

4.3 Qualitative Experimental Results

Figs. 5 and 6 show qualitative evaluation of RF-GAN where two sample images are translated for each of the seven studied datasets. For the multi-domain translation dataset, one sample is selected for each season and translated to the other seasons. We can see that RF-GAN produces good-quality translations for both cross-domain and multi-domain translation consistently across all translation tasks.

4.4 Ablation Study

An ablation study is performed over two object transfiguration datasets apple ↔ orange and horse↔zebra to show the effectiveness of the reconfigurable generator and multi-domain discriminator in RF-GAN. Two new ablation models

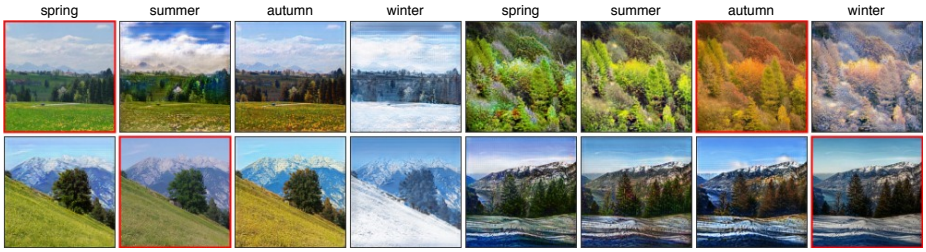

Fig. 6. Input images and translated images by RF-GAN for different seasons of the Alps. Input images highlighted by red boxes are translated to other seasons.

	CycleGAN	Abl. 1	Abl. 2	RF-GAN
# gen. / dis. / params. (M)	2 / 2 / 28.286	2 / 1 / 25.521	1 / 2 / 16.908	1 / 1 /14.143
apple→orange	181.34	177.03	175.79	173.13
orange→apple	164.46	140.05	139.96	136.64
horse→zebra	81.25	66.97	65.68	38.23
zebra→horse	143.03	151.07	149.24	143.77

Table 4. Ablation studies: Abl. 1 replaces two CycleGAN discriminators with our multi-domain discriminator. Abl. 2 replaces two CycleGAN generators with our reconfigurable generator. The results show that our reconfigurable generator and multi-domain discriminator outperform the CycleGAN generators and discriminators clearly.

Abl. 1 and Abl. 2 are trained as shown in Table 4 where Abl. 1 replaces CycleGAN's two discriminators with our multi-domain discriminator and Abl. 2 replaces CycleGAN's two generators with our reconfigurable generator. As Table 4 shows, our reconfigurable generator and multi-domain discriminator both outperform CycleGAN's generators and discriminators clearly. While combined, the complete RF-GAN produces the best FID.

5 Conclusion

This paper presents a reconfigurable GAN (RF-GAN) that is small yet capable of translating images realistically. Different from state-of-the-art translation GANs that usually have large model size and network parameters, RF-GAN learns a single reconfigurable generator that can perform bidirectional translations by swapping its parameters. In addition, RF-GAN has a multi-domain discriminator that allows bidirectional discrimination without requiring domain-specific discriminators or additional classifiers. RF-GAN reduces the model size by up to 75% as compared with state-of-the-art translation GANs, and extensive experiments over eight datasets demonstrate its superior performance in FID. We expect that reconfigurable generative networks will inspire new insights and attract more interest in translating high-fidelity images in the near future.

References

1. Chen, X., Xu, C., Yang, X., Tao, D.: Attention-gan for object transfiguration in wild images. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 164–180
2. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. CVPR (2017)
3. Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. ICML'17, JMLR.org (2017) 1857–1865
4. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R., eds.: Advances in Neural Information Processing Systems 29. Curran Associates, Inc. (2016) 469–477
5. Mejjati, Y.A., Richardt, C., Tompkin, J., Cosker, D., Kim, K.I.: Unsupervised attention-guided image-to-image translation. In: Advances in Neural Information Processing Systems. (2018) 3693–3703
6. Taigman, Y., Polyak, A., Wolf, L.: Unsupervised cross-domain image generation. ArXiv **abs/1611.02200** (2016)
7. Gatys, L.A., Bethge, M., Hertzmann, A., Shechtman, E.: Preserving color in neural artistic style transfer. arXiv preprint arXiv:1606.05897 (2016)
8. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 2414–2423
9. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision. (2016)
10. Liu, H., Navarrete Michelini, P., Zhu, D.: Artsy-gan: A style transfer system with improved quality, diversity and performance. (2018) 79–84
11. Tomei, M., Cornia, M., Baraldi, L., Cucchiara, R.: Art2real: unfolding the reality of artworks via semantically-aware image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 5849–5859
12. Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.: Texture networks: Feed-forward synthesis of textures and stylized images. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. ICML'16, JMLR.org (2016) 1349–1357
13. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, USA, Curran Associates Inc. (2017) 700–708
14. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on. (2017)
15. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2018) 8789–8797
16. Lin, J., Chen, Z., Xia, Y., Liu, S., Qin, T., Luo, J.: Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation. IEEE transactions on pattern analysis and machine intelligence (2019)
17. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Proceedings of the

27th International Conference on Neural Information Processing Systems - Volume 2. NIPS'14, Cambridge, MA, USA, MIT Press (2014) 2672–2680
18. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In Precup, D., Teh, Y.W., eds.: Proceedings of the 34th International Conference on Machine Learning. Volume 70 of Proceedings of Machine Learning Research., International Convention Centre, Sydney, Australia, PMLR (2017) 214–223
19. Denton, E., Chintala, S., Szlam, A., Fergus, R.: Deep generative image models using a laplacian pyramid of adversarial networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. NIPS'15, Cambridge, MA, USA, MIT Press (2015) 1486–1494
20. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
21. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16, USA, Curran Associates Inc. (2016) 2234–2242
22. Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., Efros, A.: Context encoders: Feature learning by inpainting. (2016)
23. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. ArXiv **abs/1809.11096** (2018)
24. Dumoulin, V., Belghazi, I., Poole, B., Lamb, A., Arjovsky, M., Mastropietro, O., Courville, A.: Adversarially learned inference. (2016)
25. Mirza, M., Osindero, S.: Conditional generative adversarial nets. ArXiv **abs/1411.1784** (2014)
26. Zhan, F., Zhu, H., Lu, S.: Spatial fusion gan for image synthesis. 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 3653–3662
27. Zhan, F., Xue, C., Lu, S.: Ga-dan: Geometry-aware domain adaptation network for scene text detection and recognition. In: Proceedings of the IEEE International Conference on Computer Vision. (2019) 9105–9115
28. Vondrick, C., Pirsiavash, H., Torralba, A.: Generating videos with scene dynamics. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16, USA, Curran Associates Inc. (2016) 613–621
29. Wu, J., Zhang, C., Xue, T., Freeman, W.T., Tenenbaum, J.B.: Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16, USA, Curran Associates Inc. (2016) 82–90
30. Anoosheh, A., Agustsson, E., Timofte, R., Van Gool, L.: Combogan: Unrestrained scalability for image domain translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (2018) 783–790
31. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). ICCV '15, Washington, DC, USA, IEEE Computer Society (2015) 2650–2658
32. Laffont, P.Y., Ren, Z., Tao, X., Qian, C., Hays, J.: Transient attributes for high-level understanding and editing of outdoor scenes. ACM Trans. Graph. **33** (2014) 149:1–149:11
33. Lee, H.Y., Tseng, H.Y., Mao, Q., Huang, J.B., Lu, Y.D., Singh, M., Yang, M.H.: Drit++: Diverse image-to-image translation via disentangled representations. arXiv preprint arXiv:1905.01270 (2019)

34. Shih, Y., Paris, S., Durand, F., Freeman, W.T.: Data-driven hallucination of different times of day from a single outdoor photo. ACM Trans. Graph. **32** (2013) 200:1–200:11
35. Wang, X., Gupta, A.: Generative image modeling using style and structure adversarial networks. ArXiv **abs/1603.05631** (2016)
36. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: Unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE international conference on computer vision. (2017) 2849–2857
37. Yu, X., Cai, X., Ying, Z., Li, T., Li, G.: Singlegan: Image-to-image translation by a single-generator network using multiple generative adversarial learning. In: Asian Conference on Computer Vision. (2018)
38. Zhang, R., Isola, P., Efros, A.: Colorful image colorization. Volume 9907. (2016) 649–666
39. Sangkloy, P., Lu, J., Fang, C., Yu, F., Hays, J.: Scribbler: Controlling deep image synthesis with sketch and color. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 6836–6845
40. Karacan, L., Akata, Z., Erdem, A., Erdem, E.: Learning to generate images of outdoor scenes from attributes and semantic layouts. CoRR **abs/1612.00215** (2016)
41. Gan, Z., Chen, L., Wang, W., Pu, Y., Zhang, Y., Liu, H., Li, C., Carin, L.: Triangle generative adversarial networks. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, USA, Curran Associates Inc. (2017) 5253–5262
42. Zhu, J.Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward multimodal image-to-image translation. In: Advances in Neural Information Processing Systems. (2017)
43. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: ECCV. (2018)
44. Lee, H.Y., Tseng, H.Y., Huang, J.B., Singh, M.K., Yang, M.H.: Diverse image-to-image translation via disentangled representations. In: European Conference on Computer Vision. (2018)
45. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770–778
46. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. (2017) 2794–2802
47. Li, C., Wand, M.: Precomputed real-time texture synthesis with markovian generative adversarial networks. In: European conference on computer vision, Springer (2016) 702–716
48. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2014)
49. Borji, A.: Pros and cons of gan evaluation measures. Computer Vision and Image Understanding **179** (2018) 41–65
50. Lucic, M., Kurach, K., Michalski, M., Gelly, S., Bousquet, O.: Are gans created equal? a large-scale study. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., eds.: Advances in Neural Information Processing Systems 31. Curran Associates, Inc. (2018) 700–709
51. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Pro-

ceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, USA, Curran Associates Inc. (2017) 6629–6640
52. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 2818–2826